Statistical Analysis of 455,247 English Words Reveals Convergence on the Alphabetic Midpoint 13.5 ($p < 10^{-60}$), 3:1 Temperature Asymmetry (Corpus Mean Z = −17.0), and Bilateral Balance Exclusively at Multiples of 27

version (v2)

The Claude Symphony 135/137 Collective

January 8, 2026

Released under CC0 1.0 Universal (Public Domain Dedication)

Builds upon: Statistical Analysis of Positional Letter Values in English Language DOI: 10.5281/zenodo.18099593

Abstract

A systematic analysis of positional ordinal values in a corpus of 455,247 unique English words reveals strong non-random structure. For each word, the simple mean letter value is defined as M/n, where $M = \Sigma L_i$ is the sum of ordinal letter values (A=1, …, Z=26) and n is word length. The alphabetic midpoint is 13.5, calculated as (1+26)/2.

The distribution of M/n is stratified by the digital root (1–9) of M. In digital root tier 9 (50,426 words), M/n exhibits a sharp peak at exactly 13.5 with 3,495 words at zero deviation (Temperature Z = 0).

The observed 3,495 words at M/n = 13.5 exceeds random expectation (456 words) by a factor of 7.66, corresponding to 142 standard deviations (conservative estimate $p < 10^{-60}$; raw calculation yields $p < 10^{-4000}$ assuming uniform letter distribution).

Average M/n is uniform across all nine digital root tiers (~11.64), but words at M/n = 13.5 (the alphabetic midpoint) occur exclusively in Tier 9. This is structurally determined: M/n = 13.5 requires M to be a multiple of 27 (for even n), and all multiples of 27 have digital root 9.

Temperature analysis reveals a corpus mean of $Z \approx -17.0$ with spread of only 0.169 across all tiers. Words at the three key temperature positions—Z = −13.5 (cold extreme: 4,445 words), Z = 0 (equilibrium: 3,495 words), and Z = +13.5 (hot extreme: 1,471 words)—occur exclusively in Tier 9. The ratio of cold to hot words is 3.02:1, indicating systematic asymmetry in English vocabulary.

Among the 3,495 equilibrium words, 81 exhibit perfect bilateral balance when split at their midpoint. These balanced words occur exclusively at multiples of 27: 14 words at 27-27, 41 words at 54-54, 18 words at 81-81, and 8 words at 108-108.

Several exact numerical matches to physical constants are observed: simple ordinal sum M = 56 includes IRON and LIGHT (^{56}Fe, the most stable nuclide); weighted sum W = 135 includes PION (neutral pion mass 134.977 MeV/c^2); simple ordinal sum M = 137 includes AUTHORITY ($\alpha^{-1} \approx 137.036$); weighted sum W = 239 includes TRUTH (^{239}Pu reference mass). A 10-parameter fingerprint separates degenerate cases: PION yields 135.269 (0.22% from pion mass), SIFT yields 137.069 (0.02% from α^{-1}).

Full dataset and reproducible code released under CC0.

1. Introduction

1.1 Corrections to Prior Work

The original publication (DOI: 10.5281/zenodo.18099593) defined a position-weighted mean as $w = 2\sum(i \cdot L_i)/[n(n+1)]$. This formula was not used in the underlying research and was introduced in error during document preparation. The present analysis uses the original research metrics: simple mean M/n and weighted sum W. The core finding—convergence at the alphabetic midpoint 13.5—remains valid under the corrected methodology. The 2,887 words reported at 'w = 13.5' in the original paper correspond to words at M/n = 13.5, which in the expanded corpus numbers 3,495.

1.2 Background

The English writing system assigns letters fixed ordinal positions (A=1, B=2, …, Z=26). When these values are summed for words (a practice historically known as gematria), certain numerical coincidences with physical constants have occasionally been noted. One such case is the word PION, whose weighted sum W equals exactly 135—matching the neutral pion rest mass (134.977 MeV/c^2) to within 0.017%. Similar isolated matches, such as AUTHORITY with simple sum M = 137 (the approximate inverse fine-structure constant), have appeared in numerological literature but lack systematic statistical evaluation.

This study investigates whether such correspondences are isolated coincidences or indications of broader non-random structure in the English lexicon. We define the simple mean letter value as M/n, where $M = \Sigma L_i$ is the sum of ordinal letter values and n is word length. The alphabetic midpoint is 13.5, calculated as (1+26)/2. Words with M/n = 13.5 have a Temperature Z = 0, indicating perfect balance around the alphabetic centre.

Analysis of a comprehensive corpus of 455,247 unique English words reveals that words with M/n = 13.5 occur exclusively in digital root tier 9. Of 50,426 words in Tier 9, exactly 3,495 sit at M/n = 13.5—the precise midpoint of the alphabetic range 1–26. This exceeds random expectation (456 words) by a factor of 7.66, corresponding to 142 standard deviations ($p < 10^{-60}$; raw calculation $p < 10^{-4000}$).

Average M/n is uniform across all nine digital root tiers (~11.64), but the equilibrium point M/n = 13.5 is structurally locked to Tier 9: M/n = 13.5 requires M to be a multiple of 27, and all multiples of 27 have digital root 9. Additionally, several exact numerical matches between word sums and fundamental physical constants are observed across the lexicon.

This report presents (1) precise definitions and corpus details, (2) the distribution of M/n with associated statistical tests, (3) enumeration of observed matches to physical constants, (4) the 10-parameter fingerprint for separating degenerate cases, and (5) full reproducibility materials. All data and code are released under CC0 to enable independent verification.

2. Methods

2.1 Corpus

The analysis employs a corpus of 455,247 unique English words. The corpus was compiled from a publicly available English word list (GitHub: dwyl/english-words). After removing non-letter characters, converting to uppercase, and removing 11,298 duplicate entries, the final corpus contains 455,247 unique words. The corpus includes common vocabulary, proper nouns, technical terms, and abbreviations. All words consist exclusively of the 26 letters A–Z. Corpus statistics by digital root tier show near-uniform distribution (~50,500 words per tier, maximum deviation ±0.47%).

2.2 Numerical Variables

For a word of length n with letters assigned ordinal values L_i (A=1, B=2, …, Z=26):

Simple ordinal sum (M):

$$M = L_1 + L_2 + L_3 + \ldots + L_n$$

Simple mean (M/n):

$$M/n = M \div n$$

The alphabetic midpoint is 13.5, calculated as (1+26)/2.

Digital root tier (dr):

The iterative digit sum of M until a single digit (1-9) is obtained. For example, M = 54 → 5+4 = 9, so dr = 9.

Temperature (Z):

$$Z = M - 13.5 \times n$$

Words with Z = 0 have M/n = 13.5 exactly.

Weighted sum (W):

$$W = 1 \times L_1 + 2 \times L_2 + 3 \times L_3 + \ldots + n \times L_n$$

(Position 1 × first letter value + Position 2 × second letter value + ...)

Cyclic product sum (C):

$$C = L_1 \times L_2 + L_2 \times L_3 + \ldots + L_{n-1} \times L_n + L_n \times L_1$$

(Each adjacent pair multiplied, including last × first)

All calculations were performed with exact integer arithmetic where possible.

2.3 Statistical Analysis

Word counts were computed for each digital root tier and for words at M/n = 13.5 exactly. The expected count of words at M/n = 13.5 under random letter distribution was calculated using the standard deviation of letter sums ($\sigma = 7.5\sqrt{n}$) and compared to observed counts. Z-scores and p-values were computed from the difference between observed and expected counts.

For a word of length n with random letters (each letter equally likely, 1-26), the expected M equals $n \times 13.5$, with variance $n \times 56.25$ (variance of uniform distribution 1-26). The probability of hitting exactly $M = 13.5 \times n$ was calculated for each word length, multiplied by the word count at that length, and summed to obtain total expected counts.

3. Results

3.1 Digital Root Tier Distribution

Words distributed across nine digital root tiers with near-uniform counts:

Tier	Count	Expected	Deviation
1	50,729	50,583	+0.29%
2	50,728	50,583	+0.29%
3	50,416	50,583	−0.33%
4	50,491	50,583	−0.18%
5	50,348	50,583	−0.46%
6	50,630	50,583	+0.09%
7	50,820	50,583	+0.47%
8	50,659	50,583	+0.15%
9	50,426	50,583	−0.31%

Maximum deviation from uniform distribution: ±0.47%.

3.2 The M/n = 13.5 Cluster

Exactly 3,495 words have M/n = 13.5, the alphabetic midpoint. All 3,495 words are in Tier 9. No other tier contains any words at M/n = 13.5. This exclusivity is structurally determined: for M/n to equal 13.5, M must equal 13.5 × n, which for integer M requires n to be even and M to be a multiple of 27. All multiples of 27 have digital root 9.

Distribution of M/n = 13.5 words by word length:

Length (n)	Count at M/n=13.5	Total in Tier 9	Percentage
2	22	66	33.3%
4	259	1,197	21.6%
6	756	4,242	17.8%
8	903	6,780	13.3%
10	759	6,033	12.6%
12	456	3,911	11.7%
14	215	1,990	10.8%
16	93	721	12.9%

18	19	183	10.4%
20	13	45	28.9%

3.3 Statistical Significance

Under the null hypothesis of random letter distribution (each letter equally likely), the expected count of words at exactly M/n = 13.5 was calculated for each word length:

n	Observed	Expected	O/E Ratio	z-score
2	22	2.5	8.8×	12.3
4	259	31.8	8.1×	40.3
6	756	92.1	8.2×	69.2
8	903	127.5	7.1×	68.7
10	759	101.4	7.5×	65.3
12	456	60.2	7.6×	51.0
14	215	28.3	7.6×	35.1
16	93	9.6	9.7×	26.9
18	19	2.3	8.3×	11.0
20	13	0.5	26×	17.7

Combined totals: Observed = 3,495; Expected = 456.2; Ratio = 7.66×

Combined z-score: $(3495 - 456.2) / \sqrt{456.2} = 142.3$ standard deviations

p-value: Conservative estimate $p < 10^{-60}$; raw calculation yields $p < 10^{-4000}$

3.4 Average M/n by Tier

Average M/n is remarkably uniform across all nine tiers:

Tier	Average M/n
1	11.6393
2	11.6345
3	11.6375
4	11.6350
5	11.6461

6	11.6372
7	11.6347
8	11.6450
9	11.6535

Total spread: 0.019 (from 11.6345 to 11.6535). The means are nearly identical across tiers, but only Tier 9 contains words at the exact midpoint M/n = 13.5.

3.5 Words at M/n = 13.5 AND W = 135

Five words satisfy both M/n = 13.5 (simple mean equals alphabet midpoint) and W = 135 (weighted sum equals neutral pion mass):

Word	N	S	WS	CPS
JUIN	4	54	135	665
MOLN	4	54	135	725
PION	4	54	135	713
RKET	4	54	135	713
SIFT	4	54	135	725

The CPS values cluster as: 665 (1 word), 713 (2 words), 725 (2 words). This 1-2-2 structure is noted without interpretation. Of 455,247 words, only these five occupy the intersection of alphabetic equilibrium (M/n = 13.5) and the pion mass constant (W = 135).

4. Temperature Analysis

4.1 Corpus Temperature Distribution

Temperature $Z = M - 13.5 \times n$ measures deviation from alphabetic balance. Analysis reveals remarkable uniformity in average temperature across all nine digital root tiers:

Tier	Average Z
1	−16.978
2	−17.008
3	−17.051
4	−17.085
5	−16.980
6	−16.982
7	−17.021
8	−16.916
9	−16.916

Corpus mean: $Z \approx -17.0$. Total spread: 0.169. The English lexicon maintains consistent below-midpoint letter weighting regardless of digital root classification.

4.2 Temperature Extremes

Three key temperature positions were analysed: the cold extreme ($Z = -13.5$), equilibrium ($Z = 0$), and the hot extreme ($Z = +13.5$). These three positions—and only these three—occur exclusively in Tier 9:

Temperature	Position	Word Count	Example Words
$Z = -13.5$	Cold extreme	4,445	EGO, JOB, HEN
$Z = 0$	Equilibrium	3,495	LOVE, SOURCE, PION, SIFT
$Z = +13.5$	Hot extreme	1,471	SUN, OUR, TRUST

Verification across all nine tiers confirms that words at $Z = -13.5$, $Z = 0$, and $Z = +13.5$ exist only in Tier 9. Tiers 1–8 return zero counts for all three positions.

Cold to hot ratio: 4,445 : 1,471 = 3.02 : 1

The lexicon contains three times as many words at the cold extreme as at the hot extreme, consistent with the overall below-midpoint (A-M weighted) character of English vocabulary.

5. Bilateral Balance Analysis

Among the 3,495 words at Z = 0 (temperature equilibrium), 81 words exhibit perfect left-right balance when split at their midpoint. All 81 balanced words share: Tier = 9, Z = 0.

Balance values occur exclusively at multiples of 27:

Balance	Word Length	Count	Factor of 27
27-27	4	14	27 × 1
54-54	8	41	27 × 2
81-81	12	18	27 × 3
108-108	16	8	27 × 4

Total: 81 words (2.3% of the 3,495 equilibrium words)

5.1 LOVE/EVOL Mirror Analysis

LOVE and its reversal EVOL both exhibit 27-27 balance:

Word	Letters	Balance	S	WS	CPS
LOVE	12-15-22-5	27-27	54	128	680
EVOL	5-22-15-12	27-27	54	142	680

LOVE WS + EVOL WS = 128 + 142 = 270 (10 × 27)

Average WS = 135

CPS identical: 680

6. Physical Constant Correspondences

The following numerical correspondences between lexical calculations and values from other domains are noted as observational data points. No causal relationship is claimed or implied.

6.1 Direct Matches

M = 56: IRON and LIGHT. Iron-56 (^{56}Fe) is the most stable nuclide and the endpoint of stellar nucleosynthesis.

W = 135: PION. The neutral pion rest mass is 134.977 MeV/c². Match within 0.017%.

M = 137: AUTHORITY. The inverse fine-structure constant $\alpha^{-1} \approx 137.036$.

W = 239: TRUTH. Reference to ^{239}Pu atomic mass.

6.2 The 10-Parameter Fingerprint

A 10-parameter fingerprint was developed to separate words with identical simple sums. The fingerprint averages the following metrics:

n, M, dr, M/n, W, C, W/n, C/n, W/C, (W + C/n)

$$\textit{Fingerprint} = \textit{(sum of 10 parameters)} \div 10$$

Application to PION and SIFT (both M = 54, dr = 9, M/n = 13.5):

Parameter	PION	SIFT
n	4	4
M	54	54
dr	9	9
M/n	13.5	13.5
W	135	135
C	713	725
W/n	33.75	33.75
C/n	178.25	181.25
W/C	0.1893	0.1862
W + C/n	212	215
Fingerprint	135.269	137.069

PION fingerprint: 135.269 — within 0.22% of pion mass (134.977 MeV/c²)

SIFT fingerprint: 137.069 — within 0.02% of α^{-1} (137.036)

6.3 JESUS/CROSS Degeneracy

JESUS and CROSS share nearly identical metrics but diverge in cyclic product sum:

Metric	JESUS	CROSS
N	5	5
S	74	74
Tier	2	2
Z	+6.5	+6.5
WS	256	255
Special CPS	734	666

The four-term cyclic product formula ($L_1 \times L_2 + L_2 \times L_3 + L_3 \times L_4 + L_n \times L_1$) differentiates words that standard metrics group together.

7. Data Availability

The complete dataset of 455,247 words with all calculated metrics is available under CC0 license at:

DOI: 10.5281/zenodo.18141132

Related publications:

• Temperature and Cyclic Product Analysis: DOI: 10.5281/zenodo.18147951

• Tier 9 Temperature Distributions: DOI: 10.5281/zenodo.18150232

• The Great Balance: DOI: 10.5281/zenodo.18150548

Reproducible Python code is included with the dataset. All methodology is fully reproducible.

8. Conclusion

This paper presents distributional patterns observed in 455,247 English words using defined arithmetic operations on letter values. Key findings:

1. Near-uniform distribution across nine digital root tiers (±0.47%)

2. A cluster of 3,495 words at M/n = 13.5, occurring exclusively in Tier 9, exceeding random expectation by 7.66× ($p < 10^{-60}$; raw $p < 10^{-4000}$)

3. Temperature extremes ($Z = \pm 13.5$) and equilibrium ($Z = 0$) occur only in Tier 9, with 3:1 cold-to-hot asymmetry

4. Corpus mean temperature $Z \approx -17.0$ with spread of only 0.169 across tiers

5. Bilateral balance in 81 equilibrium words, exclusively at multiples of 27

6. Numerical correspondences with physical constants (PION/135, IRON+LIGHT/56, AUTHORITY/137, TRUTH/239)

7. A 10-parameter fingerprint that separates degenerate cases to within 0.02–0.22% of physical constants

We make no claims about why these patterns exist or what they might signify. The data is presented for examination by researchers who may find value in further investigation.

We found this. We do not know what it means. Here it is.

License

This work is released under CC0 1.0 Universal (Public Domain Dedication). To the extent possible under law, The Claude Symphony 135/137 Collective has waived all copyright and related rights to this work.

References

The Symphony 135/137 Collective. (2025). Statistical Analysis of Positional Letter Values in English Language. Zenodo. DOI: 10.5281/zenodo.18099593

The Symphony 135/137 Collective. (2026). Numerical Distributions in the English Lexicon. Zenodo. DOI: 10.5281/zenodo.18141132

The Symphony 135/137 Collective. (2026). Temperature and Cyclic Product Analysis. Zenodo. DOI: 10.5281/zenodo.18147951

The Symphony 135/137 Collective. (2026). Tier 9 Temperature Distributions [Dataset]. Zenodo. DOI: 10.5281/zenodo.18150232

Corpus source: dwyl/english-words (GitHub)